DOCUMENT RESUME

ED 242 763 TM 840 176

AUTHOR	Estes, Gary D.; Arter, Judith A.
TITLE	Item Banking for State and Local Test Development and Use: Specific Guidelines for Local Practitioners.
INSTITUTION	Northwest Regional Educational Lab., Portland, OR. Assessment and Evaluation Program.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	[Jan 84]
CONTRACT	400-83-0005
NOTE	69p.
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Educational Objectives; Educational Planning; Elementary Secondary Education; *Guidelines; *Item Banks; Language Arts; Mathematical Concepts; Reading Skills; *School Districts; *State Departments of Education; *Test Construction

ABSTRACT

A project to assist with improving test efficiency in local school districts and state agencies is described. The report describes and defines the term "item banking." Considerations in selecting, developing, or managing an item bank are discussed. Testing purposes considered suitable for item banking include student selection for special programs, certification of student competencies, general survey assessments, and diagnosis or monitoring individual student progress on instructional objectives. Resources which local districts, state agencies, and others might provide, and roles each agency might play, are discussed. The report is designed to provide the background and guidance needed to form decisions concerning whether and how to use item banking. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM EA
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

ED242763

ITEM BANKING FOR STATE AND LOCAL TEST DEVELOPMENT AND USE[1]

Specific Guidelines for Local Practitioners

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy

Gary D. Estes and Judith A. Arter

Assessment and Evaluation Program
Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue
Portland, Oregon 97204

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
M.M. Rogers + G.D. Estes
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 176

[1]This monograph is based upon work performed pursuant to Contract 400-83-0005 of the National Institute of Education. It does not, however, necessarily reflect the views of that agency.

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
II.	ISSUES RELATED TO DEVELOPING, ADOPTING AND USING ITEM BANKS	7
III.	PRACTICAL CONSIDERATIONS IN ITEM BANKING	17
IV.	POTENTIAL ROLES FOR FEDERAL, STATE AND LOCAL AGENCIES	39
	BIBLIOGRAPHY	44
	Appendix A: Questions About Item Banking Systems	51
	Appendix B: Potential Item Information Variables	59
	Appendix C: Objectives List for NWREL	61

ITEM BANKING FOR STATE AND LOCAL TEST DEVELOPMENT AND USE

I.

INTRODUCTION

The role of accountability in local school districts and state education agencies continues to be an important concern. Tests have long been the primary benchmarks by which effective schools or education are measured. Decreases or increases in test scores are often the most public measures of school and district effectiveness. The movement toward minimum competency requirements during the late seventies is a primary example of accountability in action. Minimum competency requirements often result in districts and states developing their own tests or making available test items or resources which can be used to assess minimum competencies. Budget and cost concerns also stimulate interest in finding testing systems which can assess critical objectives for a minimum cost.

Concerns such as these provide the background which led the Northwest Regional Educational Laboratory (NWREL) to undertake a project to assist with improving test efficiency in local school districts and state agencies. NWREL, in consultation with local district and state department staff, decided that a project which addressed the potential of item banking for state and local use and which brought together the various considerations associated with item banking for local test development and use might be the most efficient focus for the project. Thus, the project was designed to answer the following questions and provide information on resources for assistance in each:

1. What are the issues and questions to address in deciding whether to use item banking and other testing options?

2. If item banking is a viable alternative, what are the issues and questions one must address in choosing an item banking approach?

3. Which item bank characteristics and issues are critical for the different testing purposes?

4. What roles and resources can state and local agencies provide to assist each other?

These questions are addressed in this paper.

The paper and project is to serve those in local school districts and state departments of education who are concerned with improving their testing programs through the use of test item banks. College and university examination services and armed services' testing programs are examples of others who might also profit from ideas presented here.

Fitting Into a Larger Context

In designing the project, we sought to merge our work with other efforts which addressed such special testing concerns as how to use standardized tests, how to improve classroom use of tests, how to select, manage and use commercial testing resources and how item banking or related applications might best be used. Those efforts have included the UCLA Center for the Study of Evaluation's study of classroom testing, Oregon Teaching Research's efforts to integrate testing and teaching, Michigan State's work on matching test content with instructional materials and University of Pittsburg's studies of microcomputer applications for assisting teachers with testing.

We sought also to integrate this project with prior NWREL studies. For instance, NWREL has conducted some development and research in the area of item banking. NWREL's Guide to Item Banking in Education (1980) provides a listing of sources for test items which might be available to state and local agencies. It also includes some discussion on item classification schemes for setting up an item bank. NWREL also examined the relative costs and outcomes

of test development under an item banking versus a traditional item writing framework (Brzezinski and Demaline, 1982). That study provided information about effects and resources for using item banks for test development.

Paper Organization

The paper is organized as follows. First, a description and definition is given for the term "item bank." This is to insure that readers will have a common referent and understanding of the terms used in the paper. Then we discuss general issues in item banking. Considerations in selecting, developing or managing an item bank are addressed. Third, we provide assistance in deciding whether one should consider item banking in a particular context or to serve a specific local or state purpose for testing. Testing purposes considered include selecting students for special programs, certifying student competencies, conducting general survey assessments, and diagnosing or monitoring individual student progress on instructional objectives or units. Fourth, resources which local districts, state agencies and others might provide and the roles each agency might play are discussed briefly. This discussion is intended to stimulate thinking about cooperative roles available to state agencies and local districts.

This paper is intended to provide background and guidance needed to make the decisions whether and how to use item banking. The relationship between those questions and sections in the paper is noted in Figure A. In short, the paper is structured around this flow of decisions.

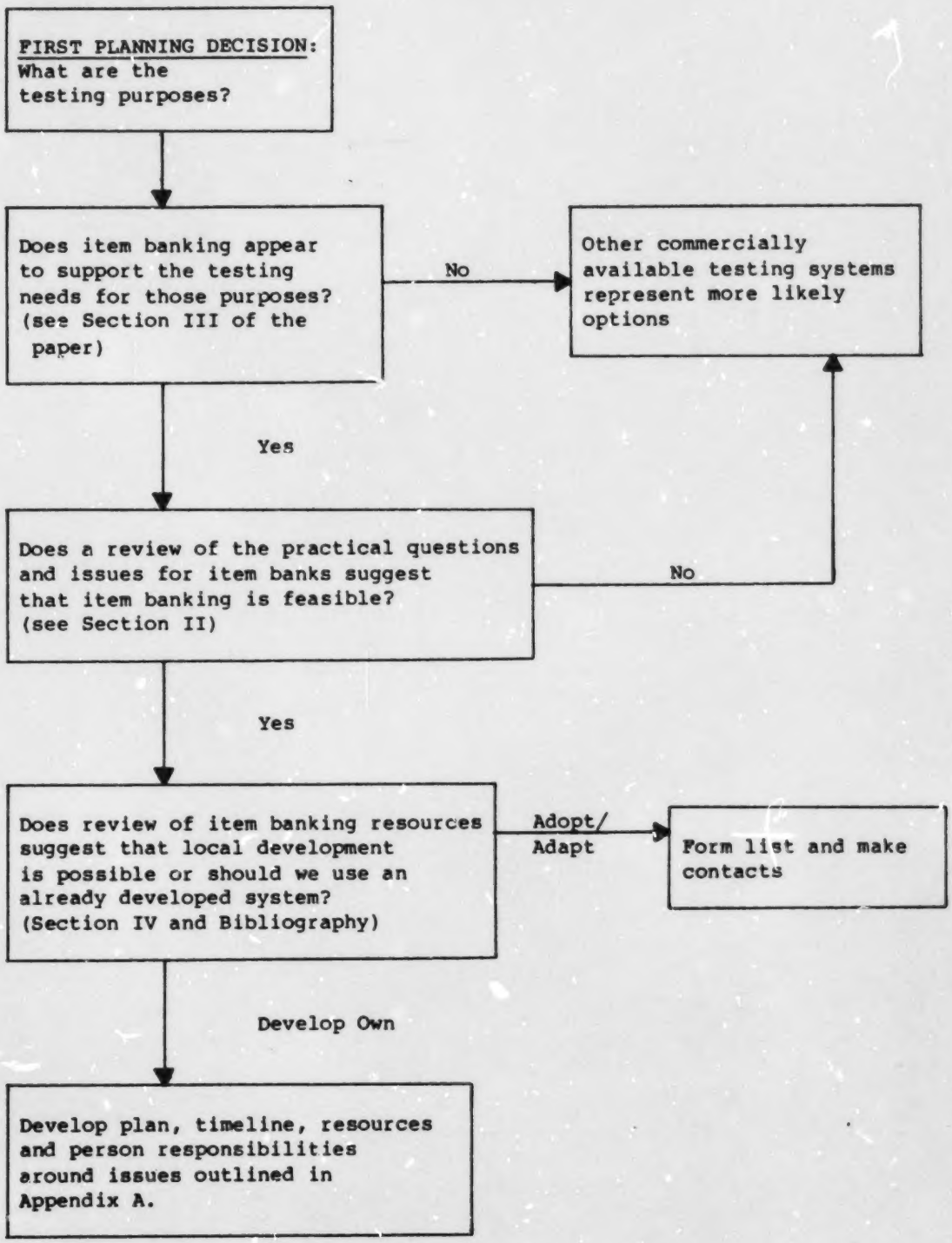

Figure 1

Flow Chart of Planning Questions to be Addressed in Designing and Using Item Banks

Item Bank Definition and Description

An "item bank" is defined differently depending upon one's interest, orientation or purpose for use (e.g., Newbould and Massey, 1977). For example, Lippey (1974) defined item banking as "the storage of questions in machine readable form." This definition is understandable, given that Lippey's interest was in computer-assisted test construction. On the other hand, those interested in aggregating data across different tests might define item banking in terms of scale or statistical requirements. For example, Mead (1981) states, "the central idea in building and maintaining an item bank is to calibrate all the items onto a common variable." From still another perspective, those interested in making instructional decisions might define item banks in terms of test classification schemes, e.g., Hiscox and Brzezinski (1980), Davis and Zacharies (1982) and Zacharies (1979). Thus, the term item banking takes on different meaning as a function of perspective.

For purposes of this paper, an item bank will be defined as a large collection of distinguishable test items. "Large" means that the number of items available is greater than the number to be used in any one testing or test. "Collection" allows for items to be developed by the user or others, but implies all are kept together in some retrievable form e.g., on computer tapes or on file in hard copy. "Distinguishable" means that the items carry some information that permits the test constructor to control the selection of items for a test. This information might be an assigned description of the item (e.g., content/objective classifications) or be a measured aspect of the item (e.g., difficulty level). This definition does not require that items be stored in the computer. Thus, issues such as computer-assisted test printing, scoring and reporting which are related to computer-assisted testing are not requirements for our item banking definition. Issues in these areas are

applicable to traditional as well as to item banking test development. Neither does our definition require that all items in an item bank necessarily represent any single content domain, be appropriate for any single population, or be in any single format such as multiple-choice, essay or performance task statements.

Webster defines a bank as "An establishment where money is received for custody and repaid on demand." Test items can also be banked. The concept of depositing for later withdrawal and for use by others to serve their particular purposes underscores the meaning of an item bank as discussed here. Our inclusive definition of an item bank allows for an examination of the issues and concerns involved in item bank applications for current educational uses of assessment. It is clear that some item bank characteristics, e.g., calibrated items, will be important for some purpose or combinations of purposes. Finally, when an item bank is to serve multiple testing purposes the benefits of calibrated items increases. Although some, e.g., Mead, 1981, stress the advantages of a calibrated item bank, we do not emphasize them in our discussion.

II.

ISSUES RELATED TO DEVELOPING, ADOPTING AND USING ITEM BANKS

When designing an item bank, there are many issues to be addressed and decisions to be made. In this section, we discuss these issues at some length. The issues were identified from a review of literature and information on item banking. Over 75 sources on item banks were reviewed for information on cautions, important design considerations, advantages and disadvantages. A list of those sources is provided in the Bibliography at the end of this paper. We have synthesized this information and organized it into five categories--issues surrounding (a) test items, (b) total tests, (c) management systems, (d) user acceptance, and (e) costs.

This discussion and the lists in Appendix A will be helpful in developing or adopting an item banking approach for a specific testing purpose. Various of the issues will be germane to different testing purposes. In this section, we will merely define the issues. We will discuss the issues in relation to testing purposes in the next section.

Before beginning, however, we must give some general cautions about these issues. First, some of the issues also apply to other testing options and, thus, are not specific to item banking. This means that any concerns implied from an issue apply to testing in general and not just item banking. Second, this discussion is only an overview of issues based on literature reviews and conversations on item banks. We plan to collect information on evaluations or reports of specific item banks to update the issues and weigh their importance. Third, the issues interrelate. For example, test assembly may depend on system issues such as available computer hardware and software. As

presented here, the issues were organized in a logical order. But that order may change, and indeed questions that are critical may change, depending on the specific item bank.

Issues Related to Test Items

The major decisions to be made concerning test items are (a) how to acquire or develop items, (b) how to classify items for later use, and (c) how to manage the items within the item bank.

Acquisition and development of items. In local test development one needs to decide whether to acquire existing items or develop items to meet local purposes. Additionally, one needs to consider the number and type of items which will be needed to meet local objectives and purposes (Dunkleburger, 1974; Reckase, 1981; Johnson, 1982; Prosser, 1974). Types of items can range from multiple-choice to essay and performance tasks. The requirements for maintaining different item parts, e.g., associated reading passages and test directions, within an item bank vary and will need to be considered. If items are developed locally, one needs to decide procedures for reviewing, evaluating and piloting items (if possible) before incorporating them into the item bank. If one plans to acquire items, sources and resources need to be considered. Existing materials need to be evaluated. Depending on the subject matter and the number of items needed, different mixes of acquisition and development would be indicated. For example, there are probably more sources for previously developed math items than science or social studies items. See Curtis, 1972 for a further discussion of whether users should write their own items.

Classifying items. Given that most local test development and item banking is at least motivated by a need to test local goals and objectives, the matching of items to local priorities--the classification of items--is critical. Items will be drawn from the bank based in large part on what they test, i.e., their classification. Twenty-five potential item information variables are contained in Appendix B. These include such characteristics as subject matter, instructional objectives, cross-referencing codes, types of item statistics, item source, last use, cross references to instructional materials, etc. It is unlikely that any given item will be classified in terms of all variables. The subset of variables used will depend upon the purpose for the bank and the local setting. We suggest that all interested parties and users be involved in the process of designing the item classification.

Typically, if one adopts a previously developed item bank, items will come with a classification system that will need to be adapted for local use. Hiscox and Brzezinski (1980) outline several classification systems. Whatever classification scheme is adopted, it should be flexible so as to accommodate new items and categories of items. For instance, NWREL has incorporated reading, language and math items from other classification systems into our framework with little difficulty.

Management of items. Procedures will need to be established for updating and revising items. Questions such as, "Who will make and approve item revisions?" and "Should revised items replace or simply be added to the item bank?" are key issues. Other item management decisions involve who will be responsible for item management, how will duplicate items be identified, how will provision be made to update the classification scheme and how will users be informed of updates. If the items are to be stored in a computer format,

procedures are needed for handling the graphics that accompany some items as well as other non-textual information or enhancements. Finally, items which use common directions, reading passages, graphs or similar information must be managed so as to allow these associated parts to be coordinated. Such material might be repeated for each item or separate systems might be established with cross-references used to link relevant parts to each item.

Test Issues

The major decisions which have to be made at the test level are (a) how the tests will be assembled; (b) what provisions will be made for administering, scoring and reporting the results; and (c) how the tests constructed from the item bank will be evaluated. Many of these decisions will depend on the extent to which a computer will be used. The need for and use of computers is discussed later in greater detail under systems considerations. We suggest, however, that in actual practice, test issues and computer-related decisions be considered together. Therefore, we related them to some extent below.

Test Assembly. Given that an item bank will have more items than will be needed for any one test, a key issue is how subsets of items will be selected. Item selection depends on (a) the item classification scheme, (b) the information which is stored on items, and (c) the characteristics of the needed test. Item banks which are not computerized and which require manually selecting, typing and editing items incur many of the same test development costs and limitations as test development projects conducted without item banks. Therefore, for users requiring fairly frequent, large scale and sophisticated test development (such as several parallel test forms each year), manual item bank systems may not be the best answer. A manual system

is more acceptable if test assembly is infrequent or when tests are short or are reused over an extended period (such as in annual screening, program evaluation or assessment efforts). Finally, manual systems which maintain "camera ready" hard copies offer some of the efficiency potential of more automated computerized systems.

In a computerized system, test development and assembly depends on what information is stored (e.g., Ansfield, 1973; Buckley-Sharp, 1973; Hazlett, 1973; Seely and Willis, 1976). Items are selected and tests are structured from specifications provided by the developer. Common specifications include test length, content and difficulty. The computer might select randomly from all items which meet the specifications or pick the most or least frequently used items. The computer can also select supporting item information such as reading passages, common directions for items and graphics. It can also be instructed to consider such technical item characteristics as item difficulty in assembling a test, if such data are stored.

Test assembly often takes the form of a requirement such as this: "Develop a 40-item reading test with four items for each of 10 objectives but no more then eight reading passages can be used in a 45-minute test." Some item banks might use only one reading passage per item whereas others might have an algorithm for selecting items and reading passages to meet the specifications outlined, i.e., finding the set of eight reading passages with accompanying items to test each of the 10 objectives with four items. Such complex assembly might require both computerized and manual test assembly to maximize test quality and test development efficiency.

Administering scoring and reporting. Once a system is designed for selecting items and a test is produced, one needs to consider administration, scoring and reporting concerns. Again, these decisions depend on the extent

to which a computer will be used. A manual item bank can be accompanied by manual scoring and reporting or by computer scoring and reporting, although provision would have to be made for entering the answer key in the latter case. Likewise, an item bank stored on computer can be supported by manual or computer-assisted assembly, administration, scoring and reporting. A sophisticated computerized system might automatically develop an answer key, score tests and print out results. The level of sophistication desired depends on frequency and type of use and on the need for fast turnaround.

One test administration alternative available with a computer is to have it present items and have students' responses be entered at the terminal or keyboard. Student answers can be scored and results presented immediately. Another more sophisticated computer application is called adaptive testing (McBride, 1977; Jensewa, 1977; Wood, 1976). Adaptive testing is most often a computerized testing procedure in which students take the test on-line, and the computer selects the next item to be administered depending on the student's answers of the previous items. Obviously, for this application, a computerized item bank with certain kinds of item statistics and software would be required. Testing until a student misses "x" items in a row for a ceiling and answers correctly "x" items in a row for a floor is a traditional, manual approximation to adaptive testing. Using multiple test forms with varying difficulties is another manual approximation.

It is clear that one of the major potential benefits of item banks is the ability to obtain test data closely related to local objectives or classification schemes. Incorporating objective and subtest scores into scoring and reporting will be keys to maximizing this potential benefit of item banks.

Evaluation. Given that an item bank affords the potential to reuse items and tests, a system needs to be considered for evaluating items included in and tests constructed from the item bank. Item performance can be monitored to detect items to revise or discard. This monitoring can range from user critiques to statistical analyses. Tests should be monitored with respect to questions of test reliability and validity for various purposes as well as appropriateness of norms. Item banks can account for these factors.

System Issues

"System" refers to the entire process for supporting item and test level activities. In other words, the manual procedures and any software or hardware for maintaining and using an item bank include (a) designing, planning and developing the system; (b) acquiring or developing manual systems, computer software and hardware; (c) training staff and monitoring system use; and (d) adopting access and security policies.

Planning and development issues. Once the decisions about methods for handling item and test issues are made, the issues for system requirements to accommodate these decisions will be clearer. Decisions and issues from this area include the persons responsible for acquiring, developing and maintaining the system, the sources needed to develop and implement the system, and whether the system is only for local use or is to be transportable to others. While the initial concern in local test development will not place high priority on transportability, it is possible that systems designed to be transportable can facilitate use and cost sharing for development and maintenance.

Software and hardware features. It will be important to decide which aspects of test development and use will be computerized and which will be manual. As described before, these decisions should be considered in conjunction with purposes, item and test considerations. Decisions need to be made as to what will be computerized: item codes that are used to refer to items stored elsewhere, item text, enhancements or other information about items such as sources or statistics. Further, it will be important to decide whether separate files are maintained for items, item information, directions and enhancements, and how the files cross-reference each other. Decisions in these areas will then result in greater clarity about hardware requirements to maintain the system. The most desirable configuration can be balanced against existing hardware capabilities and development costs for minimum needs. Many state and local districts have computer systems which already support some testing functions, e.g., scoring and reporting tests, and which might support other functions such as storing and retrieving items in the test assembly stages. Costs to develop and maintain a computer system should be compared to costs for a manual system. Also to be considered is whether system features are matched to intended users. For example, on-line access would require a "friendly system" (Itowze, 1978; Dudley, 1973).

Monitoring and training. Users will need to be familiar with the system features. Methods for training and updating users will be needed. Finally, as stated earlier, a system for monitoring item and test performance will be needed to insure that accumulated information is incorporated into subsequent efforts.

Access and security. Questions of access to the system--who and how--are key to the item bank system. Decisions in this area depend on how the systems are to be used. For example, if the system is to support mastery learning or

diagnosis, access needs to be at the classroom level. Allowing users direct access accommodates individual needs and schedules. If the use is developing equivalent forms for district or state minimum competency tests, restricted access might be critical to maintain test security. Access also can be either from photocopies of items or through direct interaction with the computer, i.e., on-line. The number of users and need for quick response time will also influence access decisions.

Use and Acceptance

No testing scheme will be accepted if it is not useful (Dunkelberger and Heikkinen, 1982; Seely and Willis, 1976). In addressing utility, several practical concerns must be considered. Utility can be maximized if the designer knows in advance what features users want, what actually will be used, who to consult in the design phase, how to establish credibility, what turnaround time is needed and what resources are necessary and available to develop tests using items from the item bank. As we have stated before, systems designed with a clear sense of testing purposes are systems that will be used.

Key aspects often overlooked in test development are the steps needed to reference test items and information to curriculum materials and instructional practices used by those who will use the tests. These steps are critical to facilitate use and acceptance. Different testing purposes have different needs for cross-referencing test items to instructional materials and strategies. Just writing items referenced to objectives does not insure that alignment will occur between curriculum, instruction and assessment. Recent information on effective school practices stresses the importance of alignment and coordination of these components in educational programs. Tests must obviously line up with both objectives and curriculum materials and practices.

Cost

The two most obvious reasons for developing an item bank for local test development are (a) to reduce testing costs or implement more efficient testing systems and (b) to obtain a testing system which is highly relevant to local objectives and needs (see, e.g., Prosser, 1973). Unless system development costs are very low, it is apparent that cost-efficiency can only be achieved with a high degree of item bank use. It is important to assess the relative costs of item banking approaches with other test development approaches given local purposes. It will be helpful to compare the costs of purchasing commercially available tests with the costs of such alternatives as (a) securing an item bank from other public agencies. This might be done by purchasing or leasing the item bank or trading for other information, (b) adapting an item bank developed by another agency or organization, (c) using or purchasing an externally developed item bank, such as those of commercial test developers to construct tests from objectives and other test specifications lists, or (d) developing an item bank by organizing and adding to the existing locally developed items and securing or developing new or additional items as needed.

III.

PRACTICAL CONSIDERATIONS IN ITEM BANKING

The relative advantages and disadvantages of an item bank depend on the application--the purpose for the testing effort, how an item bank would be configured, the resources available for the testing effort and the local climate for testing (Brown, 1973; Dunkleberger and Heikkinen, 1983; Chelu and Elton, 1977; Zacharias, 1979; Johnson and Mahr, 1983). In some situations and for some testing purposes, item banks would have more disadvantages than advantages. An example might be a local district assessment program which occurs once a year and is designed only to get an overall view of how students compare to some skill standard or some norm group. In this case, the most viable alternative for testing might be a published standardized test.

Further, different situations and testing purposes might take advantage of different features of item banks. For example, a minimum competency testing program might benefit from an item bank which provided for secure tests and items and comparable test forms for repeated testing. In contrast, a mastery or individualized learning program might benefit from an item bank which facilitates assembling tests matched to individual student objectives with minimum concerns about item or test security or comparability across test forms.

The purpose of this section is to discuss item bank design features which are more critical for various applications.

In general, the types of situations in which item banks have the most potential are when:

1. existing tests are not acceptable and the climate or staff require that tests be constructed according to local specifications.

2. frequent testing is required, as in mastery learning and individual educational programs.

3. one needs different subsets of items such as when short tests tailored to students' achievement are used in testing.

4. multiple, parallel forms of a test are required to facilitate repeated and secure testing as in minimum competency testing.

5. testing requires items which are difficult or expensive to write, such as reading passages or items using graphics or artwork.

In general, the situations in which item banking would probably be less viable are when:

1. the local climate or staff testing needs are met with existing tests.

2. the purposes for testing do not require frequent, parallel testing of the same content or rigid test security.

3. item writing and test construction is relatively easy, i.e., does not include complex item writing. It is generally easier to write math computation items than it is to write a good reading passage.

4. resources are not available to support the effort.

In short, the decision to pursue an item bank application depends primarily on the testing context and need. Cost, time and resources will also affect the applicability of item banking. In other words, if money, time, equipment and expertise are available to develop and maintain a testing system to accomplish a specific purpose, options such as item banking, published standardized tests, and locally developed tests must be compared to determine which are more cost efficient.

To assist in this comparison process, we will consider various common testing purposes in terms of (a) the requirements for tests to satisfy this purpose and (b) the relative utility, efficiency and validity of item banks versus other testing options in addressing these requirements. For purposes of this discussion, we will use the following definitions.

Utility. Utility refers to the ability of the testing scheme to satisfy the testing requirements for a given testing purpose. For example, a minimum competency test might need to assess relevant objectives and have secure, parallel forms. Various testing possibilities (such as item banks and published standardized tests) might lend themselves more or less to these needs. We will assess the utility of item banks for various testing purposes by the degree to which the testing purpose requires local specification of test content, frequent testing, multiple parallel forms of a test, more items than are needed for any one test and items which are difficult or more expensive to write. If a given testing application does not require one or more of these features, item banking becomes less of a useful alternative.

Efficiency. Testing options vary in their relative efficiency, i.e., costs versus results. Different testing applications require various levels and types of financial resources, time and expertise for both development and maintenance. For example, item banking can require a large amount of initial and ongoing resources when compared with using existing tests. To determine whether item banking is worthwhile, one must weigh the alternatives in terms of their initial and ongoing costs. In general, a testing option will be the most efficient when it is the least costly way to achieve a testing goal. For item banking, this efficiency generally occurs when multiple forms of a test tied directly to instruction require locally, individually constructed tests. Item bank efficiency will derive from the availability of existing items organized for easy access in test development, assembly and administration.

Technical Quality. High quality tests are valid and reliable for a given purpose. Item banks have a potential to increase the quality of locally developed tests because they (a) often include only items of proven quality, i.e., have been pilot tested and reviewed, (b) offer increased test

reliability by maximizing the match of tests with students' abilities, (c) increase content validity due to better match of test and instructional program content, and (d) increase test security (Ansfield, 1973; Brown, 1973). Item banks will generally have an advantage over published tests in terms of content validity but not necessarily other technical quality. Item banks will generally have an advantage over locally made tests, i.e., local item writing to construct tests, in terms of technical quality but not necessarily content validity. Therefore, the advantage of an item bank over other testing options depends on the relative need for various technical features and the relative need to precisely match the curriculum.

Testing Purposes and Item Banking

A testing option has the most potential in a given application when it provides an efficient option for producing high quality tests which will be most useful for a given testing function. A summary of the utility, efficiency and quality characteristics of item banks for various testing purposes is summarized in Tables 1 and 2.

Table 1 summarizes the needs which an item bank might fulfill for seven common testing purposes. The five needs are those situations in which an item bank is most frequently applicable. A "high" is a favorable assessment for an item bank's potential as compared to other options. A "low" implies that the need which the item bank might fulfill is not relevant for that testing purpose, e.g., a district or statewide survey assessment does not require individually tailored or frequent testing. Table 2 provides a summary of ratings of utility, efficiency and quality for each testing purpose. A "high" rating, again, implies that item banking compares favorably or has potential

for more utility, efficiency or quality than other options. A "low" rating implies that an item bank will be potentially less useful, less efficient or have less quality than other options. Other primary options are local test development without an item bank or using externally developed tests.

The ratings in Tables 1 and 2 will not apply or be consistent across all contexts or even all individuals within a specific context and purpose. One's rating will depend in part on the individual's values. The relative ratings for each area or criterion across purposes is probably more dependable than the assignment of low or high. Our purposes for including the ratings are to highlight areas where item banking might be more fruitful. Finally, using a testing system for multiple purposes will result in a higher benefit for item banking. Multiple purposes will also increase item bank requirements, e.g., the need for calibrated items.

Table 1

Summary of the Advantages of Item Banking
for Testing Purposes

Areas In Which Item Banks Have the Most Benefit / Testing Purpose	Student Screening	Classroom Testing	MCT Minimum Competency Testing	Student Diagnosis	Mastery Learning	Survey Assessment	Program Evaluation
1. Need to match local objectives	Low/Medium	High	High	High	High	Medium	Medium
2. Need for frequent testing	Low	High	Medium	Medium/High	High	Low	Low
3. Need for individually tailored tests, e.g., diagnosis, adaptive testing	Low/Medium	Low	Low	Medium	Medium	Low	Low
4. Need for multiple, equivalent forms for secure or repeated testing	Low	Low	Hign	Low	High	Low	Medium/Low
5. Need for assistance with more difficult items, e.g., reading passages, graphs	*	*	*	*	*	*	*

*Depends on the area being tested.

Table 2

Utility, Efficiency and Technical Quality of Item Banking
Compared to Other Testing Options

Criteria \ Testing Purpose	Student Screening/ Placement	Classroom Testing	MCT Minimum Competency Testing	Student Diagnosis	Mastery Learning	Survey Assessment	Program Evaluation
Utility: will be more useful and used than "non-item bank" tests, e.g., commercial	Low/Medium	High	High	High	High	Low	Low/Medium
Efficiency: will be less costly than either: (a) purchase test (b) develop own test	Low	High	Medium	Medium	High	Low	Low
Technical Quality: will be more valid, e.g., (a) instructional/content validity	a. Medium	High	a. High: Instructional	a. High	a. High if adaptive type testing	a. Medium	a. High
(b) and external/ predictive validity	b. Low		b. Low-Medium	b. High	b. Medium	b. Low	b. Low

Testing Purpose: Screening and Placement of Students

Description. Local programs, either school based or district-wide, often need to screen and place large numbers of students in special programs such as special education, compensatory education or gifted programs. Some sort of testing is usually involved in this process, although observational checklists and teacher recommendations are also often used as part of the process.

Testing Needs. For student screening and placement, one needs to rank or group students so that the lowest or highest achieving students can be identified. Scores must be available for individual students. Specific match of test content and curriculum is not necessarily required because the criterion for screening and placement into major educational programs are usually based on general achievement and ability indicators. It is only necessary that there are enough items and enough range in the difficulty among items to rank order students in a reliable and valid manner. Strictly parallel forms of a test are not necessarily needed since students are generally tested for this purpose and only once a year or less often. Security is also not of high importance since the same tests might not be needed repeatedly for the same students, i.e., screening is usually not more than once a year, nor will actual placement be based on a single test score. Of course, as screening approximates final student placements, test security will be an issue.

Utility, Efficiency and Validity of Item Banks For This Application. Item banks would not have high efficiency or utility advantages, if screening and placement were the only use. This application does not require frequent testing, more items than are needed for any one test, or multiple parallel forms of the same test. The need for match to local content is less of a concern for this than for some other applications. In general, items for

screening are not difficult to write. An exception to this might be in the area of oral language where the need is to screen for language dominance or proficiency. With respect to the validity of the test, item banks do not offer any significant technical advantage over other alternatives. Content validity would probably improve somewhat, but might not be needed.

Other Options. Other options would be annual testing or assessment activities which are already in place. Since screening generally occurs yearly, information from these regularly occurring activities should be available on a routine basis. Typical examples of such yearly activities are achievement testing, regular determination of skills mastered and minimum competency testing. Teacher screening checklists could also be used. For other applications, e.g., special education, commercial diagnostic tests might suffice. One of these options would probably offer the same content and technical validity as an item bank and would probably be more efficient in terms of time, cost and expertise.

Critical Item Banking Issues. If item banking is considered for the purpose of large scale screening, one should consider several design features. Acquire items where possible and only then develop items to fill in the gaps to minimize test development costs. A fairly general classification scheme in terms of grade level and subject matter should suffice. Item calibrations or other scale scores are probably not critical since strictly parallel forms do not necessarily need to be developed, and because students generally are tested once each year or even less frequently for screening. Since frequent testing is not needed, a manual system of item access, test assembly, scoring and reporting would probably suffice. If the screening

system is state mandated and tests are provided by the state then a state level manual system might suffice. If the state is providing tests to local districts based on their own needs, a more carefully cross-referenced system would be needed and would probably require a computer to enter, store and retrieve items. Tests and system security are probably not critical. Use and acceptance of the system can be obtained through orientation and explanation to teachers and others. Developmental costs should be minimized in most instances except, perhaps oral language.

Testing Purpose: Classroom Testing

Description. Teachers frequently use tests in the classroom to monitor progress in an area of instruction or to produce classroom grades.

Testing Needs. For classroom testing one needs to be able to generate test items which match the content taught. Having content covered is more important than having items with a range of difficulty. Scores must be available for individual students. Security for any one test is only a short-term concern.

Utility, Efficiency and Validity of Item Banks for This Application. This application can require frequent testing and the need to match specific content. Therefore, item banking has a definite application in this area (Chelv and Elta, 1977; SOCRATES).

Other Options. Other options for testing are those currently being used--teacher-made tests or tests embedded in various curriculum series. Teacher-made tests can fall victim to test quality problems. Embedded tests might not have the flexibility required in all situations. Also, not all published materials have embedded tests, and not all teachers use standard materials.

Critical Item Banking Issues. If an item bank is to be used for this application, it should allow for direct teacher access, fast turn-around time, e.g., on-line, lots of items, comprehensive coverage of all content domains of interest, a good classification scheme, flexibility in terms of generating, storing and revising tests and automatic generation of answer keys. One approach is to begin entering items on a locally available microcomputer for repeated use and retrieval.

Testing Purpose: Minimum Competency Testing

Description. Minimum competency testing involves making decisions about whether an individual has acquired the minimum skills necessary for some purpose. For example, certain skills might be required to be eligible for high school graduation, to be promoted to the next grade or to be certified to perform some job. The focus is on whether the examinee has mastered a defined set of skills and not on the examinee's full range of knowledge or skills.

Testing Needs. Very important educational decisions are often made on the basis of these tests. The test must focus on those skills or objectives considered minimally necessary. The pass/fail criterion must accurately differentiate between those who are competent and those who are not. Individual scores need to be available for each student. Frequent testing is likely and the need for test security is high. Multiple equivalent forms are needed for the frequent testing and security. Finally, reading passages, graphs or similar difficult items can lesson the effort to secure items closely related to local objectives.

Utility, Efficiency and Validity of Item Banking for This Application. Item banks have the potential to be useful and efficient for minimum competency testing. Although minimum competency testing generally occurs only once or twice a year at some grades, equivalent forms may be needed for retesting those who initially fail; the content must clearly reflect local objectives; and there is a great need for test security. These needs can all be met by having an item bank with a large number of items which are carefully cross-referenced and carry item statistics such as item calibrations. Item banking offers advantages in the area of technical quality, since achieving good quality and local content validity may be more easily achieved with item banking than with other approaches.

Other Testing Options. States and districts currently involved in minimum competency testing often choose other testing options. The two most widespread choices are locally developed tests and published standardized tests. In both cases the test is matched to the minimum objectives and a cut-off is established. There are disadvantages to both of these approaches. First, with locally developed tests, although the content validity is good, much effort needs to be expended for each test to provide technical quality. If test forms need to be changed or developed, the overall effort becomes burdensome. With published tests, although the technical quality of items is good, the match to the local curriculum may not be sufficiently close. Therefore, there can be questions as to the validity of using published tests for this purpose.

Critical Item Bank Issues. Items must carefully measure those objectives outlined for minimum competence. Therefore, the classification scheme needs to directly reflect the minimum competency skills. Security might dictate different tests each year. Therefore, the number of items per objective

should be large. Items should be carefully screened for technical adequacy. Since equivalent forms need to be equated, items need to carry item statistics. Access to the items for updating, adding and deleting will have to be carefully controlled. Access would depend on whether the process is state or locally controlled. Since the test specifications will change infrequently, a standard assembly process could be developed. Since the content is fixed, the item bank need only cover those objectives to be tested. This limits the size requirements of the bank. Since tests are generally assembled only once a year, or less often, the production, scoring and reporting could be either computer assisted or manual depending on the number of students to be processed. All interested parties need to be carefully and thoroughly informed as to how the item bank relates to the minimum competency testing program.

Testing Purpose: Diagnosis of Student Skills

Description. This application involves assessing the skills and knowledge the student has mastered so that that appropriate instruction can be provided. Diagnosis can occur on a broad scale at the beginning of the school year, at designated times during the year for special populations, or as the need arises in the classroom for individual students.

Testing Needs. Since diagnosis is used to plan instruction, the test content needs to closely match instructional priorities. There needs to be enough items per objective so that an accurate estimate of ability on that objective can be obtained. Scores are needed for individual students. Frequency of testing depends on the nature of the curriculum and the types of students served. The system should be flexible enough to handle any frequency or volume of testing in a variety of areas. Very frequent and routine

assessment embedded in the curriculum puts the effort more in the realm of mastery learning (see the next section). Diagnosis does not, however, necessarily need to support a mastery learning setting. For example, when a student in a program seems to be having problems, the teacher might use a test to identify the skill areas rather than to monitor movement across skills. Equivalent test forms are not particularly critical since repeated testing will fall into mastery assessment more than the diagnosis function. Tests are controlled at the classroom level and there is no major security concern. The existence of well written reading passages, pictures and charts will be helpful.

Utility, Efficiency and Validity of Item Banking for This Application. Item banking has a strong potential for supporting diagnosis. Increase in the number of potential users, need for frequent testing and variability in the subsets of items needed will result in an item bank becoming more viable and useful than other choices. Diagnosis requires detailed and varied content specifications and short turnaround time. The largest obstacle to usefulness in this area is user acceptance. Many teachers have their preferred methods and tests for diagnosis. Provisions for incorporating multiple sources of items and updating users will insure the system has high utility. The number of test forms and degree to which local objectives are not assessed by available tests will increase the efficiency and technical quality, particularly the content match. The extent to which the item bank extends and organizes existing tests and procedures will also insure its efficiency.

Other Testing Options. Other testing options include published standardized tests, curriculum embedded tests which might accompany the instructional materials or teacher-made tests. Each of these would be a more or less viable alternative depending on the circumstances. For infrequent

testing which is used for broad level placement into materials, a published standardized test might suffice. As frequency and variability increase, curriculum embedded and locally developed tests might be better. With still greater frequency and variability, an item bank might be best.

Critical Item Bank Issues. Because of the emphasis on curriculum match, an item bank needs to have a classification scheme which enables users to examine many areas, levels of difficulty and subskills. Therefore, the emphasis is more on the nature of the classification scheme than on item statistics. If the classification scheme is useful, the bank will be successful. If testing is infrequent and used more to place students into major groupings, access to the item bank for test development could be more centralized. If testing is more frequent and used to readjust instruction during the course of the year, test development access should be more at the classroom or teacher level. This has implications for where the item bank is located and the need for remote terminals if the system is computerized. Provision should be made for easy access for purposes of updating the items. If maintained at the class or school level, the bank could be manual. With increasing users and number of items a manual system could become awkward. An item bank could be developed by having teachers submit any locally developed tests items currently in use. Establishing the organization and classification schemes are critical steps in this process. Provision would need to be made for informing and updating teachers and users.

Purpose: Mastery Learning

Description. A mastery learning approach generally involves having students progress, at their own rate, through a curriculum that has a detailed scope and sequence. Mastery of each objective is tested before the student goes to the next objective.

Testing Needs. The testing scheme should allow students to be tested on the same content at different times and should allow for retesting of the same students who did not pass a previous test. Thus, the system should allow for many unique tests of the same content. Tests need to be available at the classroom level, development and scoring need to have quick turnaround, each student's place in the curriculum needs to be known and test administration schedules should be flexible. This purpose involves potentially frequent testing on specific local classroom instructional objectives. Item banks can aid with this need and with developing multiple forms. Individually tailored tests and assistance from reading passes, etc., can benefit also but are not as critical as frequent, well matched tests.

Utility, Efficiency and Validity of Item Banking for This Application. (Hambleton, 1979; Brown, 1982; Dunkleberger and Heikkinen, 1982; Cohen and Cohen, 1973). Item banking has a major potential in this area because of the need for frequent testing, parallel forms, specific content matching and security. Although the initial system development would be time consuming and costly, if it is used as intended, the system would ultimately increase efficiency of testing and record keeping. In fact, many classes which might be interested in mastery learning do not use this approach at this time because of the record keeping burden.

Other Testing Options. Curriculum embedded tests or locally developed tests can be used depending on the circumstances. When a commercially

produced curriculum in an area is being used, their existing embedded tests might be the best alternative. For more locally developed curricula, objective tests might be "standardized" and locally developed. The disadvantage to these systems is that there is no flexibility in the test items or in the areas tested. In addition, all records generally need to be kept manually. As the intended mastery learning situation becomes more complex the advantage of item banking increases.

Critical Item Bank Issues. The classification scheme should be detailed enough and the item coverage broad enough to test any objective or set of objectives in the curriculum. Since the purpose is oriented toward instructional management, the classification scheme for items is more important than availability of item statistics. In fact, the usefulness of the item bank will depend on the classification scheme. Provision would need to be made for on-going updating and additions to the system to cover areas not previously covered adequately. Access to the items needs to be at the classroom level and turnaround for test development and scoring must be very rapid. This implies that either items will be stored in each area of use, e.g., classrooms, or a system which is computer supported will be needed. If the system is computer assisted, automatic records could be kept of where each student is in the curriculum. Testing could even be done on-line. This would decrease the paperwork burden. If the system contained enough items, they could even be used for drill and practice. The sophistication of such a system is substantial, but similar systems have been established.

Use Area: Survey Assessment

Description. Survey assessment usually occurs once a year and has the major purpose of estimating current status of students and trends in this status over time. Survey assessment can be used as a broad indicator of how well instruction is going, how populations are changing, and to help allocate resources. Survey assessment has traditionally been done in the basic skills areas of reading, math and language arts. Some areas are expanding their assessment programs to cover other areas such as science and social studies.

Minimum Testing Needs. Survey assessment usually occurs on a broad scale and emphasizes the status and progress of groups of students rather than individuals. The test should cover the broad content in areas of interest and should allow for the desired comparisons, i.e., either comparisons with a national group, comparisons between districts or states, comparisons with individual sets of objectives or comparisons across time. Test items need to measure a range of content across a range of difficulties so that accurate status can be observed. If forms and levels are to change over time, some mechanism must exist for comparing results.

Utility, Efficiency and Validity of Item Banking for This Application. Item banking has limited application to survey assessment depending on the application (see Johnson and Maher, 1982 for a discussion of national assessment). Testing is usually not frequent, multiple parallel forms are usually not needed, testing usually does not occur in areas for which items are difficult to write, and there is little need for tailored testing. There is some need to match local objectives. Item banking becomes more attractive in the situations where a closer match to local scope and sequence is desired, different forms of the test need to be developed each year, and local control over production is desired.

Other Testing Options. Locally developed tests and published standardized tests are other options currently in use. Locally developed tests are generally used when published tests are not seen to adequately measure the local curriculum. Published tests are typically used when comparisons with the national population are desired. Another option, typically underutilized, is to request tailored tests from publishers by submitting local objectives to be tested and having publishers use their item bank to generate a test. One disadvantage to this process can be turn-around time. These alternatives are generally more attractive than local maintenance of an item bank because of the ability to obtain the content and technical quality needed for this application without having the work of assembling an item bank.

Critical Item Bank Issues. If an item bank is to be used to support assessment, items need to be classified to meet the intended use. For example, items might be classified by grade level and major objectives. Since the same content is to be covered each year, the item bank need only have multiple items for those areas to be tested. Depending on the scale of the testing, test printing, scoring and reporting could be either computer assisted or manual. Test items would need to be carefully screened for quality.

Use Area: Program Evaluation

Description. Tests are often used to assess the impact of instruction on students. Rather than evaluating or making decisions about individual students, program evaluation generally focuses on results across students in a program, school, grade, or district.

Testing Needs. Testing is usually yearly although other designs are certainly acceptable and used. The emphasis is usually on pre-post-test design group averages rather than individual scores. The tests need to cover the content taught in a more or less specific manner depending on the purpose of the evaluation. If the program evaluation is to evaluate the efficiency of instruction on local objectives, then a close match is needed. If the evaluation objective is in relation to national norms, the local objectives are less critical. Parallel forms are not needed unless repeated testing is needed over a short period of time. Testing occurs at the building or district level depending on the question to be addressed. If the program evaluation will have critical decisions, i.e., program discontinuation or rewards, then test security can become critical. When local objectives or local test development is desired, then having item parts available such as reading passages can aid test development.

Utility, Efficiency and Validity of Item Banks for This Application. Item banking does not have a high utility in this area. Testing is not usually frequent, multiple parallel forms are generally not needed, nor is there a major concern for test security. There is a need for test content to match the curriculum to varying degrees depending on the evaluation criterion. Item banking is best for program evaluations in which the main goal is to measure progress on local objectives. An item bank will increase the probability that results will be used due to the actual or perceived instructional content validity.

Other Testing Options. Other tests used for program evaluation include published standardized tests and locally developed tests. Published standardized tests can be used for many evaluation situations and are particularly useful in situations in which one de res to use the norms of the test as the comparison group.

Critical Item Bank Issues. If an item bank is to be used for program evaluation, it would need to have the flexibility to address any number of content questions. Therefore, the classification scheme would have to be general and flexible. Turnaround for test development and reporting does not need to be immediate. Items need to be of high quality and item statistics need to be available. Often normative data are needed. Access to the items would have to be restricted.

How to Proceed

Although we have attempted to outline the most frequent reasons for testing and the possible applications of item banking to these purposes based on the most typical situations, each possible application needs to be carefully assessed on its own. Idiosyncratic situations could affect the decision whether to develop an item bank, or the possible applications and design of a bank. Examples of such idiosyncracies are having a staff member who is interested in item banking and is willing to spend a good deal of time on it, having equipment and software already available, desiring to blend multiple testing purposes, or having an existing pool of items from which to start. Therefore, before any decision can be made concerning item banking, one should:

- Decide on the local needs for testing. Why is testing being done? What is the minimum testing design needed to accomplish these purposes? What extras would be nice to have? These plans should be based on an assessment of the testing needs of potential users.

- Catalog resources in terms of money, personnel, time, equipment and expertise available for testing and test development. This catalog would help to outline what options are feasible.

- Consult the "Questions About Item Banking" checklist. Decide what options and concerns are important for your various applications. Balance this against resources. Do the same for other testing options.

The previous discussion and our assessments in Tables 1 and 2 offer that item banking's potential varies according to one's testing purpose. It is also reasonable to consider accomplishing several purposes with an item bank to either reduce testing burden or make testing more efficient. We do not propose that our ratings are appropriate for everyone in every context nor do we present evidence that an item bank cannot serve more than one testing purpose. We do want to be clear that combining purposes will result in tensions and tradeoffs. If one attempts to use an item bank for a minimum competency test and for mastery learning, tension will exist between maintaining the security for competency testing and having readily available tests for mastery learning. With the advent of increased technology and resources, i.e., items and software, blending purposes will be increasingly possible.

IV.

POTENTIAL ROLES FOR FEDERAL, STATE AND LOCAL AGENCIES

This section will outline and discuss some roles which agencies at the federal, state and local levels might play to assist with test development using item banks.

Federal Role

One federal role is to provide research and development support to further knowledge. This project is an example of that role. The specific need this project is responding to is to improve the ability of state and local agencies to undertake test development activities in an efficient and useful manner. A second related federal role is the development and maintenance of item banks. The primary example of this is the National Assessment of Educational Progress (NAEP), previously maintained and operated by the Education Commission of the States but recently transferred to the Educational Testing Service.

How Have You Used National Assessment Materials?... Responses From Six Educators (1982) provides a summary of ways NAEP resources can assist state and local agency test development activities. They include: NAEP objectives, test items and services. Each will be mentioned briefly.

NAEP objective booklets were used by local agencies to (a) write their own local goals and objectives, (b) compare their goals to the nationally-recognized goals in NAEP materials, and (c) update their objectives.

Released NAEP test items were also useful to these agencies. The items were well developed and field tested and were referenced to nationally recognized goals and content areas. The statistical data on the NAEP items

and related demographic data were useful for comparing district results with NAEP results. Items were also available in areas in which relatively few items exist. These areas include science and social studies, as well as other areas such as assessments of citizenship and energy awareness.

The NAEP system was helpful to districts in that it provided a model for testing and data collection. The system for differentiating skills at different skill and age levels was cited as a benefit. Direct contact and assistance from NAEP staff was cited as another benefit.

In summary, benefits available to state and local agencies included reductions in test development costs due to the availability of established objective frameworks, test items, and data for making local and national comparisons. In this 1982 report, one district pointed out that, without NAEP a $250 assessment project would have cost $5,000. The purpose of this section is not to endorse NAEP as a system which will assist with all local or state assessment needs. It is to inform readers of the availability of NAEP resources which might improve the quality and efficiency of local test development efforts as a means of illustrating a positive federal impact on testing through item banking.

State Roles

Primary roles state agencies have played in the area of test development include sharing of items or item banks and providing assistance to districts in their states. In a recent informal survey of state assessment directors, directors from 20 states indicated that they had developed or acquired test items that they could share with other state agencies. Many of these test items are in the traditional areas of reading and math and are primarily in the elementary rather than high school grades. However, items were also

reported available in less common areas such as writing, citizenship, nutrition, physical education/fitness, health, science, etc. The test items range from those which are part of state assessment programs to items for developing minimum competency tests.

State agencies have assisted districts in their states by developing objective frameworks and items for local district use. Alaska has been developing objectives and items for use by districts for over six years. California responded to their districts' minimum competency testing needs by developing items for districts to use in developing their own tests and by providing technical assistance to the districts. Wisconsin and Minnesota have also made contributions of items and item banks for local district use. The purpose of citing individual states is not to present them as necessarily the best examples nor to ignore contributions that other states have made to local test development efforts through item banks. It is to highlight that actual cases do exist where:

1. State agencies have developed items and item banks for local district use. Many states routinely use their own selected or developed items in statewide assessments.
2. State agencies have provided technical and consultation assistance to districts in their local test development efforts. This assistance has included informing the districts of available item banks and sources, assisting the districts with designing a system and providing fiscal resources for developing an item bank system which might be useful to the developing districts as well as others.

Such state roles are to be encouraged and supported.

Local Roles

A primary role for local agencies is to share information on their efforts in developing or using item banks. A number of ESEA Title III or IVC projects have developed objective or criterion referenced testing systems as part of an instructional program. The extent to which these have large number of items for repeated testing is an example of item banking potential, e.g., Stark County, Ohio and Los Angeles County, California. Other districts might benefit from similar efforts. An example of another major item bank effort in which districts shared information and resources is the Northwest Evaluation Association item bank. Essentially, over a period of several years a tri-county set of goals and objectives were developed which were subsequently made the basis for developing Rasch calibrated items to test those goals. The item banks were developed and subsequently shared with other districts in Oregon as well as outside the State, e.g., San Jose District in California.

The purpose of highlighting the sharing nature among districts is to stress that items can be used across districts and that sharing development and maintenance costs can result in more efficient systems.

Summary

This section has provided some ideas on ways federal, state and local agencies can cooperate and support local test development efforts. Cases are cited to support that:

1. NAEP test items, objectives and services can be useful to local test development and interpretations.
2. State agencies can use item banks for their state assessments and can assist local district test development through developing or making available item banks.

3. Districts can benefit from cooperative efforts in which test and item bank development costs and efforts are shared.

Finally, this section is intended to complement the previous sections by providing a guide to highlight potential sources for assistance once it appears that item banking might be a desirable alternative based on the reviews and questions posed in the previous sections. Using existing resources and contacting others will reduce the burden of a local test development effort and will improve the assessments as information is gained from previous work.

Information not included in this paper but which is planned for subsequent products include:

1. A listing and description of promising and accessible sources from which local districts and state agencies might get items, item bank systems, etc.
2. References and descriptions for microcomputer software which can be used specifically for item banking or local test development efforts.
3. References and examples of specific testing programs and ways item banking can be or is integrated into those programs.

BIBLIOGRAPHY

Ansfield, Paul J. A user oriented computing procedure for compiling and generating examinations. Ed Tech 13 #3, pp 12-13, Mar 1973. ERIC No. EJ075560.

Arneklev, Bruce L. Scaling Objective-Based Curriculum, 17 Apr 1981, 26 p. Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981). ERIC No. ED204398.

Benbasat, Izak & Others. A Database System for Course Administration. Computers and Education 6, #4, pp 333-39, 1982. ERIC No. EJ270076.

Brown, Willard A. A computer exam composition for the IBM 360/40. Ed Tech 13, #3, 1973, pp 15-16. ERIC No. EJ075562.

Brown, B. Automated test and quiz production. Classroom Computer News 2, 1982, pp 33-35.

Brzezinski, E. & Demaline, R. An empirical investigation of the results of two different test development strategies. Northwest Regional Educational Laboratory, Portland, Oregon, January, 1982.

Buckly-Sharp, M.D. A multiple choice question banking system. Ed Tech 13 #3, 1973, pp 16-18. ERIC No. EJ075563.

Byrne, Colin J. Computerized Question Banking Systems: I--the State of the Art. British Journal of Educational Technology 7 #2, pp 44-64, May 1976. ERIC No. EJ143265.

Carland & Carland. A microcomputer examination support system. A paper presented at the annual meeting of the Association for Educational Data Systems, Portland, Oregon, May 1983.

Chelu, C.J. & Elton, L.R.B. An item bank for multiple-choice questions. Physics' Education 12 #4, pp 263-7, May 1977. ERIC No. EJ179734.

Childs, Roy. Item Banking. Basic Testing Series. National Foundation for Education Research in England and Wales Publishing Co., Windsor, Berkshire, (England), 14 p. ERIC No. ED163017.

Cohen, P.S. & Cohen, L.P. Computer generated tests for a student paced course. Ed Tech 13 #3, 1973, pp 18-19.

Curtis, H.A., Ed. The Development and Management of Banks of Performance Based Test Items. Symposium presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, 1972. Harcourt Brace Jovanovich, Inc., New York, N.Y., April 1972, 36 p. ERIC No. ED072099.

David, S. & Zacharies, J. Testing in the service of education. Classroom Computer News 2, 1982, pp 28-32.

Denny, C. Question pool management system (QPMS). Ed Tech 13 #3, 1973 pp 19-20.

Dewitt, Louis J. & Weiss, David J. A Computer Software System for Adaptive Ability Measurement. Minnesota University, Minneapolis. Department of Psychology, Jan 1974, 76 p. Sponsoring Agency: Office of Naval Research, Washington, D.C., Personnel and Training Research Programs Office. ERIC No. ED088928.

Dow, Ian I. & Others. The Development of a Curriculum Change Indicator. Ontario Department of Education, Toronoto, 1978, 368 p. ERIC No. ED166831.

Dudley, T.J. How the computer assists in pacing and testing student's progress. Ed Tech 13 #3, 1973, pp 21-22.

Dunkleberger, Gary E. & Heikkinen, Henry. A review of computer-generated repeatable testing. AEDS Journal 15 #4, pp 218-25, Sum 1982. ERIC No. EJ270072.

Educational Testing Service, Princeton, N.J. Computer Assisted Assembly of Tests of Educational Testing Service, Oct. 1974, 14 p. Paper presented at the Computer Assisted Test Construction Conference (San Diego, California, October 14-16, 1974). ERIC No. ED110503.

Eisenberg, Eric M. & Book, Cassandra L. Applying Latent Trait Theory to a Course Examination System: Administration, Maintenance, and Training. Michigan State University, East Lansing. Department of Communication, 1980, 24 p. ERIC No. ED193266.

Epstein, Marion G. Classification Scheme for Items in CAAT. Educational Testing Service, Princeton, N.J., Oct. 1974, 10 p. Paper presented at the Computer Assisted Test Construction Conference (San Diego, California, October 14-16, 1974). ERIC No. ED110505.

Epstein, M.G. Computer-assisted assembly of tests at ETS. Ed Tech 13 #3, 1973, pp 23-24.

Estes, G., Millman, J. & Arter, J. Item banking for assessment in Oregon. Northwest Regional Educational Laboratory, Portland, July 1983.

Forbes, Dean W. & Ingebo, George S. An Empirical Test of the Content Homogeneity Assumption Involved in Rasch Item Calibration, April 1975, 9 p. Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975). ERIC No. ED107675.

Foreman, Dale I. SCORE: A Flexible System for School Curriculum Objective-Referenced Evaluation, April 1975, p. 10. Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975). ERIC No. ED107168.

Fuchs, Lynn & Others. Effects of Varying Item Domain and Sample Duration on Technical Characteristics of Daily Measures in Reading. Minnesota University, Minneapolis. Institute for Research on Learning Disabilities, Jan. 1981, 44 p. Sponsoring Agency: Office of Special Education (ED), Washington, D.C. ERIC No. ED211606.

Fuhs, F. Paul RIBYT--A Data Base System for Formal Testing and Self-Assessment, 1980, 10 p. ERIC No. ED193324.

Geisinger, Kurt F. A Systems Approach to Item Production and Review in a Computer Managed Instruction Project, April 1976, 20 p. Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, California, April 19-23, 1976). ERIC No. ED121280.

Hambleton, Ronald K. & Stetz, Frank P. The development of objective-based instructional programs in career education. Journal of Career Education 5 #3, pp 220-25, March 1979. ERIC No. EJ205912.

Haney, Walt & Strenio, Andrew. Summary of the Spring 1979 Conference of the National Consortium on Testing, (April 30-May 1, 1979) Huron Inst., Cambridge, Mass., 1 June 1979, 83 p. ERIC No. ED181027.

Harris, Chester W. Techniques for Analyzing Test Response Data, April 1975, 8 p. Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975). ERIC No. ED110496.

Hazlett, C.B. MEDSIRCH: Multiple choice test items. Ed Tech 13 #3, 1973, pp 24-25.

Heines, Jesse M. An Examination of the Literature on Criterion-Referenced and Computer-Assisted Testing. EDRS, November 1975, 45 p. ERIC No. ED116633.

Hiscox, M. A balance sheet for educational item bankers. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, April 1983.

Hiscox, M. & Brzezinski, E. A guide to item banking in education. Northwest Regional Educational Laboratory, Portland, Oregon, June 1980.

Howze, Glen. An interactive software system for computer-assisted testing. AEDS Journal 11 #2, pp 31-7, Winter 1978. ERIC No. EJ173608.

Hsu, T. & Carlson, A. Test construction aspects of the computer-assisted testing model. Ed Tech 13 #3, 1973, pp 26-27.

Hsu, T. & Nitko, A. Microcomputer testing software teachers can use. Paper presented at the 1983 ECS Large-Scale Assessment Conference, June. Boulder, Colorado.

Jensema, Carl J. Bayesian tailored testing and the influence of item bank characteristics. Applied Psychological Measurement 1 #1, pp 11-20, Winter 1977. ERIC No. EJ161401.

Jensen, D.D. Toward efficient, effective and humane instruction in large classes. Ed Tech 13 #3, 1973, pp 28-29.

Johnson, K. Jeffrey & Others. Computer-assisted test construction in chemistry. J. of Chemical Education 58, 1981, (February), pp 177-181.

Johnson, Sandra & Mahar, Brian. Monitoring science performance using a computerized question banking system. British Journal of Educational Technology 13 #2, pp 97-106, May 1982. ERIC No. EJ268605.

Legg, Sue M. The Use of Precalibrated Item Bank to Establish and Maintain Cutoff Scores: A Case Study of the Florida Teacher Certification Examination. March 1982, 14 p. Paper presented at the Annual Meeting of the National Council of Measurement in Education (New York, NY, March 20-22, 1982). ERIC No. ED221570.

Lewis, Meg, Villiers, Allan & Rudnick, Betty. Use of the QUEST test management system in a multi-campus environment. Paper delivered at the annual meeting of the Association for Educational Data Systems (AEDS), St. Louis, April 1980.

Lewy, Arieh & Doron, Rina. Group Tailored Tests and Some Problems of their Utilization, June 1977, 16 p. Paper presented at the International Symposium on Educational Testing (3rd, Leyden, The Netherlands, June, 1977). ERIC No. ED159210.

Libaw, F.B. Constructing tests with the MENTREXT tutorial testing system. Ed Tech 13 #3, 1973, pp 30-31.

Lippey, Gerald. Improvements to a computer-assisted test construction system. Ed Tech 16 #12, 1976, pp 26-29. ERIC No. EJ153936.

Lippey, Gerald, Ed. Computer-Assisted Test Construction. Educational Technology Publications, Englewood Cliffs, N.J., 1974, 244 p. ERIC No. ED096948.

Lippey, Gerald. Computer-Assisted Test Construction. International Business Machines Corp., San Jose, California, 1974, 2 p. ERIC No. ED095S07.

Lippey, Gerald. A Comprehensive Question Retrieval Application to Serve Classroom Teachers. International Business Machies Corp., San Jose, California, 1973, 14 p. ERIC No. ED095908.

McBride, James R. Some properties of a Bayesian adaptive ability testing strategy. Applied Psychological Measurement 1 #1, pp 121-40, Winter 1977. ERIC No. EJ161402.

McKinley, Robert L. & Reckase, Mark D. A Comparison of Procedures for Constructing Large Item Pools. Missouri University, Columbia. Tailored Testing Research Lab, August 1981, 65 p. Sponsoring Agency: Office of Naval Research, Arlington, Virginia. Personnel and Training Research Programs Office. ERIC No. ED207998.

McLean, L.D. & Ragsdale, R.G. The Rasch model for achievement tests--inappropriate in the past, inappropriate today, inappropriate tomorrow. Canadian Journal of Education 8, 1983, pp 71-76.

Mead, Ronald J. Basic Ideas in Item Banking, 14 April 1981, 19 p. Paper presented at the Annual Meeting of the National Council of Measurement in Education (Los Angeles, CA, April 14-16, 1981). Sponsoring Agency: National Board of Medical Examiners, Philadelphia, PA.; National Institute of Education (ED), Washington, D.C. ERIC No. ED208029.

Menne, J.W. Computer assisted test assembly at Iowa State University. Ed Tech 13 #3, 1973, pp 31-32.

Meyer, Carol. Feasibility study of using item banks within Oregon statewide assessment. Oregon Department of Education, June 1978.

Michener, R. Dean & Others. Automating Exams for a Statistics Course: II. A Case Study. May 1978, 40 p. Paper presented at the Annual Convention of the Association for Educational Data Systems (16th, Atlanta, GA, May 16-19, 1978). ERIC No. ED200619.

Millman, J. & Arter, J.A. Issues in Item Banking. Journal of Educational Measurement (in press).

Millman, Jason & Ourlaw, W. Scott. Testing by Computer. AEDS Journal 11 #3, pp 57-72, Spring 1978. ERIC No. EJ177688.

Muiznieks, Viktors & Dennis, J. Richard. A Look at Computer-Assisted Testing Operations. Illinois University, Urbana. Department of Secondary Education. The Illinois Series on Educational Application of Computers, No. 12e, 1979, 48 p. ERIC No. ED183191.

Newbould, C.A. & Massey, A.J. A Computerized Item Banking System (CIBS). British Journal of Educational Technology 8 #2, pp 114-23, May 1977. ERIC No. EJ179455.

O'Reilly, R.P., Gordh, W.P. & Pinsley, P. Computer-assisted test construction: an effort based on an evaluation methodology CAM system. Ed Tech 13 #3, 1973, pp 32-34.

Prosser, F. Repeatable tests. Ed Tech 13 #3, 1973, pp 34-35.

Reckase, Mark D. Tailored Testing, Measurement Problems and Latent Trait Theory, April 1981, 17 p. Paper presented at the Annual Meeting of the National Council on Measurement in Education (Los Angeles, CA, April 14-16, 1981). Sponsoring Agency: Office of Naval Research, Arlington, VA. Personnel and Training Research Programs Office. ERIC No. ED207987.

Ree, Malcom James. Automated Test Item Banking. Final Report for Period January 1977-February 1978. Air Force Human Resources Lab, Brooks AFB, Texas, May 1978, 18 p. ERIC No. ED160618.

Robitaille, David F. & Others. Mathematics Achievement Test Project. Technical Manual. British Columbia Department of Education, Victoria, February, 1980, 57 p. ERIC No. ED187531.

Ross, John A. & Maynes, Florence J. Experimental Problem-Solving Item Pool: Junior and Intermediate Division Science. Projecty #80-4029, Final R&D Report to the Ontario Institute for Studies in Education. Ontario Institute for Studies in Education, Peterborough. Trent Valley Centre, 1980, 255 p. ERIC No. ED211568.

Seely, Oliver, Jr. & Willis, Van. SOCRATES' test retrieval at the California State University and Colleges. AEDS Journal 9 #3, pp 65-70, Spring 1976. ERIC No. EJ139384.

Seely, Over, Jr. & Willis, Van. SOCRATES' Test Retrieval at the California State University and Colleges, June 1975, 5 p. Paper presented at the Conference on Computers in the Undergraduate Curricula (6th, Forth Worth, Texas, June 16-18, 1975). ERIC No. ED112828.

Tall, Graham. The possible danges of applying Rasch to school examinations and objective testing. Mathematics in School 8 #3, pp 19-22, May 1979. ERIC No. EJ205386.

Toggenburger, F. Classroom techer support system, LAUSD. Ed Tech 13 #3, 1973, pp 42-43.

Weiss, David J. Computerized Ability Testing, 1972-1975. Final Report. Minnesota University, Minneapolis, Department of Psychology, April 1976, 31 p. Sponsoring Agency: Office of Naval Research, Washington, D.C. Personnel and Training Research Programs Office. ERIC No. ED126131.

Wiggers, Thomas B. A Computerized Test Question Storage and Examination Preparation System. Educational Computer Magazine 2 #2, pp 30-31, March-Arpil, 1982. ERIC No. EJ268628.

Wood, R. Adaptive testing: a Bayesian procedure for the efficient measurement of ability. PLET 13 1976, pp 36-48.

Wolmut, Peter & Thomas, Gregory. COURSE: Computer Retrieval for Local District Planning and Evaluation, April 1981, 18 p. Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981). ERIC No. ED202875.

Yen, Wendy M. The extent, causes and importance of context effects on item parameters for two latent trait models. Journal of Educational Measurement 17 #4, pp 297-311, Winter, 1980. ERIC No. EJ239607.

Zacharias, Jerrold R. The people should get what they need: An optimistic view of what tests can be. National Elementary Principal 58 #2, pp 48-51, January 1979. ERIC No. EJ199429.

APPENDIX A

From: Millman & Arter, 1983

Questions to be Answered in Designing Item Banking Systems

I. ITEMS

A. Acquisition and Development

1. Develop/use your own item collection or use collections of others?

   a. If develop your own item collection, what development procedures will be followed?

   b. If use collections of others, will the items be leased or purchased, and is the classifcation scheme sufficiently documented and the item format specifications sufficiently compatible for easy transfer and use?

2. What types of "items" will be permitted?

   a. Will open-ended (constructed response) items, opinion questions, instructional objectives, or descriptions of performance tasks be included in the bank?

   b. Will all the items be made to fit a common format (e.g., all multiple-choice with options a, b, c, and d)?

   c. Must the items be calibrated, validated, or otherwise carry additional information?

3. What will be the size of the item collection?

   a. How many items per objective/subtopic (collection depth)?

   b. How many different topics (collection breadth)?

4. What review, tryout and editing procedures will be used?

   a. Who will perform the review/editing?

   b. Will there be a field tryout, and if so, what statistics will be gathered, and what criteria will be used for inclusion into the bank?

B. Classification

1. How will the subject matter classifications be conducted?

   a. Will the classification by subject matter use fixed categories, keywords, or some combination of the two?

   b. Who will be responsible for preparing the taxonomy?

c. How detailed will the taxonomy be? Will it be hierarchically or nonhierarchically arranged?

d. Who will assign classification indices to each item, and how will this assignment be verified?

2. What other assigned information about the items will be stored in the item bank? (See the attached list for potential attributes.)

3. What measured information about the items will be stored in the bank? (See the Appendix B list for potential measures.) How will the item measures be calculated?*

C. Management

1. Will provision be made for updating the classification scheme and items? If so:

a. Who will be permitted to make additions, deletions, and revisions?

b. What review procedures will be followed?

c. How will the changes be disseminated?

d. How will duplicate (or near duplicate) items be detected and eliminated?

e. When will a revision of an item be trivial enough that item statistics from a previous version can be aggregated with revisions from the current version?

f. Will item statistics be stored from each use, last use, or aggregated across uses?

2. How will items that require pictures, graphs, special characters, or other types of enhanced printing be handled?

3. How will items that must accompany other items, such as a series of questions about the same reading passage, be handled?

---

*This question is the subject of considerable controversy and discussion in the technical measurement literature. For example, to obtain a latent trait difficulty parameter, concern has been expressed about sample size, calibration procedure (Rasch, 3-parameter), linking models (major axis, least squares, maximum likelihood), and number of items common to the equating forms.

II. TESTS

A. Assembly

1. Must the test constructor specify the specific items to appear on the test or will the items be selected by the computer?

2. If the items are selected by the computer:

   a. How will one item out of several that matches the search specification be selected (randomly, time since last usage, frequency of previous use)?

   b. What happens if no item meets the search specifications?

   c. Will a test constructor have the option to reject a selected item, and if so, what will be the mechanism for doing so?

   d. What precautions will be taken to insure that examiners who are tested more than once do not receive the same items?

3. What item or test parameters can be specified for test assembly (item format restrictions, limits on difficulty levels, expected score distribution, expected test reliability, etc.)?

4. What assembly procedures will be available (options to multiple-choice items placed in random order, the test items placed in random order, different items on each test)?

5. Will the system print tests or just specify which items to use? If the former, how will the tests be printed or duplicated and where will the answers be displayed?

B. Administration, Scoring and Reporting

1. Will the system be capable of on-line test administration? If so:

   a. How will access be managed?

   b. Will test administration be adaptive, and if so, using what procedures?

2. Will the system provide for test scoring? If so:

   a. What scoring formula will be used (rights only, correction for guessing, partial credit for some answers, weighting by discrimination values)?

   b. How will constructed responses be evaluated (off-line by the instructor, on-line/off-line by examiners comparing their answers to a key, on-line by computer with/without employing a spelling algorithm)?

3. Will the system provide for test reporting? If so:

   a. What records will be kept (the tests themselves, individual student item responses, individual student test scores, school or other group scores) and for how long? Will new scores for individuals and groups supplement or replace old scores?

   b. What reporting options (content/format) will be available?

   c. To whom will the reports be sent?

C. Evaluation

1. Will reliability and validity data be collected? If so, what data will be collected by whom, and how will they be used?

2. Will norms be made available and, if so, based on what norm-referenced measures?

III. SYSTEM

A. Acquisition and Development

1. Who will be resonsible for acquisition/development, given what resources, and operating under what constraints?

2. Will the system be made transportable to others? What levels and what degree of documentation will be available?

B. Software/Hardware Features

1. What aspects of the system will be computer assisted?

   a. Where will the items be stored (computer, paper, card file)?

   b. Will requests be filled using a batch, on-line, or manual mode?

2. Will items be stored as one large collection or will separate files be maintained for each user?

3. How will the item banking system be constructed (from scratch; by piecing together word processing, data-base management, and other general purpose programs; by adopting existing item banking systems)?

4. What specific equipment will be needed (for storage, retrieval, interactions with the system, etc.)?

5. How user and maintenance friendly will the equipment and support programs be?

6. Who will be responsible for equipment maintenance?

C. Monitoring and Training

1. What system features will be monitored (number of items per classification category, usage by user group, number of revisions until a user is satisfied, distribution of test lengths or other test characteristics, etc.)

2. Who will monitor the system, train users, and give support (initially, ongoing)?

3. How will information about changes in system procedures be disseminated?

D. Access and Security

1. Who will have access to the items and other information in the bank (authors/owners, teachers, students)? Who can request tests?

2. Will users have direct access to the system or must they go through an intermediary?

3. What procedures will be followed to secure the contents of the item bank (if they are to be secure)?

4. Where will the contents of the item bank be housed (centrally or will each user also have a copy)?

5. Who will have access to score reports?

IV. USE AND ACCEPTANCE

A. General

1. Who decides to what uses the item bank will be put? And will these uses be the ones that the test users need and want?

2. Who will develop the tests and who will be allowed to use the system? Will these people be acceptable to the examinees and recipients of the test information?

3. Will the system be able to handle the expected demand for use?

4. Will the output of the system likely to be used and used as intended?

5. How will user acceptance and item bank credibility be enhanced?

B. Instructional Improvement. If this is an intended use:

1. Will the item bank be part of a larger instructional/decision-making system?

2. Which textbooks, curriculum guidelines, and other materials, if any, will be keyed to the bank's items? Who will make that decision and how will the assignments be validated?

3. Will items be available for drill and practice as well as for testing?

4. Will information be available to users that will assist in the diagnosis of educational needs?

C. Adaptive Testing. If this is an option:

1. How will the scheduling of the test administrations take place?

2. How will the items be selected to insure testing efficiency yet maintain content representation and avoid duplication between successive test administrations?

3. What criteria will be used to terminate testing?

4. What scoring procedures will be followed?

D. Certification of Competence. If this is an intended use:

1. Will the item bank contain measures that cover all the important component skills of the competence being assessed?

2. How many attempts at passing the test will be allowed; when? How will these attempts be monitored?

E. Program/Curriculum Evaluation. If this is an intended use

1. Will it be possible to implement the system so as to provide reliable measures of student achievement in a large number of specific performance areas?

2. Will the item bank contain measures that cover all the important stated objectives of the curriculum? That go beyond the stated objectives of the curriculum?

3. Will the item bank yield commensurable data that permit valid comparisons over time?

F. Testing and Reporting Requirements Imposed by External Agencies. If meeting these requirements is an intended use:

1. Will the system be able to handle requirements for program evaluation (e.g., Chapter 1), student selection into specially funded programs, assessing educational needs, and reporting?

2. Will the system be able to accomodate minor modifications in the testing and reporting requirements?

V. COSTS

A. Cost Feasibility

1. What are the (fixed, variable) costs (financial, time, space, equipment and supplies) to create and support the system?

2. Are these costs affordable?

B. Cost Comparisons

1. How do the item banking system costs compare to the present or other testing systems that achieve the same goals?

2. Do any expanded capabilities justify the extra cost? Are any restricted capabilities balanced by cost savings?

APPENDIX B

POTENTIAL ITEM INFORMATION VARIABLES

Assigned Item Descriptions

1. Identification number
2. Text
3. Correct/model answer; responses to incorrect answer choices
4. Required enhancements (graphs, e.g.)
5. Cross-reference to other items or to common stem (reading passage, e.g.)
6. Question type/response format (essay, true-false, e.g.)
7. Type of directions required
8. Link to previous versions
9. Source
10. Keywords or other subject matter classifiers
11. Classifier; editor
12. Thinking process (knowledge, application, e.g.)
13. Setting, illustration, or context in which the item is posed (abstract/concrete; patient situation; etc.)
14. Judged importance (essential, desirable knowledge, e.g.)
15. Recommended grade/educational level
16. Cross-reference to textbooks, curriculum guides, or other related materials
17. Security level
18. User comments

Measured Item Characteristics*

1. Index of difficulty level†
2. Index of discrimination
3. Response frequency data for each item option
4. Index of bias†
5. Average time to complete†
6. History of use (number of students tested, when last administered, frequency of use, I.D. number of students who were administered the item)
7. Readability level†

---

*Many of these item data could be recorded separately for different examinee subgroups, for separate test administrations, for serial location of the item in the test, etc.

†This item characteristic sometimes is estimated (judged) rather than measured, in which case it would appear as an assigned item description.

APPENDIX C

READING

A. COMPREHENSION AND ANALYSIS

1. Select the MAIN IDEA of a reading passage
2. Select the BEST TITLE for a reading passage
3. Identify SUPPORTING DETAILS in a reading passage
4. Identify the SEQUENCE OF EVENTS in a paragraph
5. Identify the CAUSE AND EFFECT relationship between elements in a paragraph
6. Select the correct CLASSIFICATION FOR a DESIGNATED FACT
7. Select the statement that correctly COMPARES what, when, where, why or how events happened
8. Select the statement that correctly CONTRASTS what, where, why, when or how events happened
9. Select the WORD to which a given REFERENT (pronoun, adjective or adverb) refers
10. Select the CONCLUSION given in a passage
11. Select a statement or FACT WHICH SUPPORTS the CONCLUSION
12. Select the EMOTIONAL SENSE to which the author is appealing
13. Select the PHYSICAL SENSE (e.g. sight, sound, taste, etc.) to which the author is appealing
14. IDENTIFY the FIGURE OF SPEECH in a passage
15. Identify the MEANING OF a FIGURE OF SPEECH
16. Select the statement which indicates that an INFERENCE can be drawn
17. Select meaning of GRAPHIC CLUE
18. Select the ORGANIZATIONAL PATTERN used by the author
19. Classify a statement as SPECIFIC OR A GENERALIZATION
20. Identify AUTHOR'S VIEWPOINT, bias or objectivity in a reading passage

Assessment and Measurement Program
Northwest Regional Educational Laboratory

MATHEMATICS

A. CALCULATOR ARITHMETIC

1. Compute ADDITION problems INVOLVING WHOLE NUMBERS
2. Compute ADDITION problems INVOLVING DECIMALS
3. Compute SUBTRACTION problems INVOLVING WHOLE NUMBERS
4. Compute SUBTRACTION problems INVOLVING DECIMALS
5. Compute DIVISION problems INVOLVING WHOLE NUMBERS
6. Compute DIVISION problems INVOLVING DECIMALS
7. Compute MULTIPLICATION problems INVOLVING WHOLE NUMBERS
8. Compute MULTIPLICATION problems INVOLVING DECIMALS
9. Compute POWERS
10. Compute SQUARE ROOTS
11. Perform MULTI-STEP COMPUTATIONS

Assessment and Measurement Program
Northwest Regional Educational Laboratory

LANGUAGE ARTS

A. CAPITALIZATION

1. Capitalize the FIRST WORD IN SENTENCE
2. Capitalize correctly the SALUTATION OF a LETTER
3. Capitalize correctly the CLOSING OF a LETTER
4. Capitalize the FIRST WORD IN QUOTATION
5. Capitalize correctly INTERRUPTED QUOTATIONS
6. Capitalize the PRONOUN "I"
7. Capitalize the name of the DAY OF WEEK
8. Capitalize the MONTH OF YEAR
9. Capitalize HOLIDAYS
10. Capitalize PROPER NOUNS
11. Capitalize PROPER ADJECTIVES
12. Capitalize correctly ADDRESSES
13. Capitalize correctly the TITLE OF PUBLICATIONS, MOVIES

Assessment and Measurement Program
Northwest Regional Educational Laboratory